

# Data Visualization (CIS 568/DSC 530)

Amir Akhavan

Lecture 6 (Chapter 3.1-3.5)

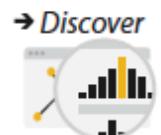
Wednesday September 16, 2020

# Why?

## Actions

### >Analyze

→ Consume



→ Discover



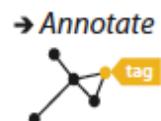
→ Present



→ Enjoy



→ Produce



→ Annotate



→ Record



→ Derive



### Search

	Target known	Target unknown
Location known	•.. •.. <i>Lookup</i>	•.. •.. <i>Browse</i>
Location unknown	← → <i>Locate</i>	← → <i>Explore</i>

### Query

→ Identify



→ Compare



→ Summarize



## Targets

### All Data

→ Trends



→ Outliers



→ Features

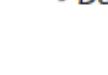


### Attributes

→ One



→ Many



→ Dependency



→ Correlation



→ Similarity



### Network Data

→ Topology



→ Paths

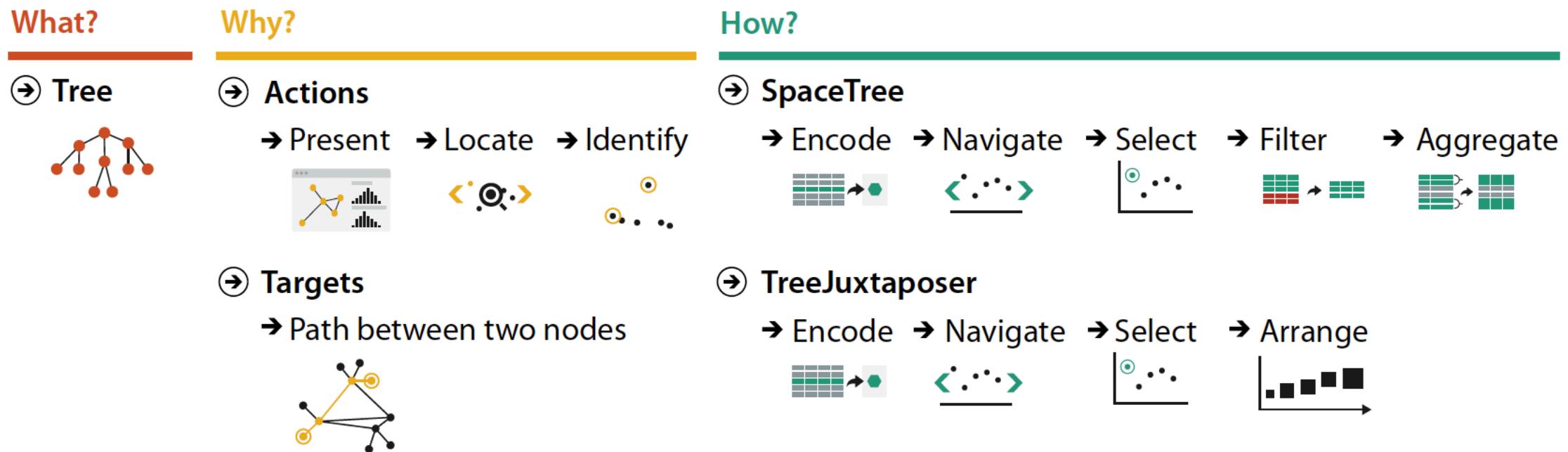


### Spatial Data

→ Shape



# The process is What, Why and How?



# Analyzing Tasks Abstractly

- Different discipline name their tasks using specific terminology  
Most of the tasks although named differently are the same

Example: immunologists and biologists call comparison of values between two groups “contrast” versus “match up”.

- One important reason to analyze the task is to understand whether and how to transform the user’s original data into different forms by deriving new data.

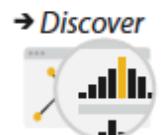
That is, the *task abstraction* should guide the *data abstraction*

# Why?

## Actions

### >Analyze

→ Consume



→ Discover



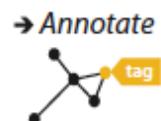
→ Present



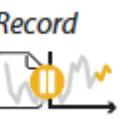
→ Enjoy



→ Produce



→ Annotate



→ Record



→ Derive



### Search

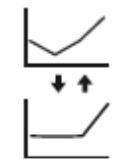
	Target known	Target unknown
Location known	•.. •.. <i>Lookup</i>	•.. •.. <i>Browse</i>
Location unknown	← → <i>Locate</i>	← → <i>Explore</i>

### Query

→ Identify



→ Compare



→ Summarize



## Targets

### All Data

→ Trends



→ Outliers



→ Features

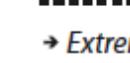


### Attributes

→ One



→ Many



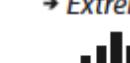
→ Distribution



→ Dependency



→ Correlation



→ Similarity



### Network Data

→ Topology



→ Paths



### Spatial Data

→ Shape



**Why?**

**Actions**

→ Analyze

→ Consume

→ Discover      → Present      → Enjoy

→ Search

→ Query

→ Summarize

→ Targets

→ All Data

→ Trends

→ Outliers

→ Features

→ Attributes

→ One

→ Many

→ Network Data

→ Paths

→ Spatial Data

→ Shape

What?

Why?

**discover** -> find new knowledge that was not previously known.

**present** -> for telling a **story** with data, or guiding an audience through a series of cognitive operations.

**Enjoy** -> not driven by a previously pressing need to verify or generate a hypothesis

Produce -> user wants to generate new material

Actions

Targets

→ Analyze  
→ Consume

→ Produce

→ Annotate



→ Record



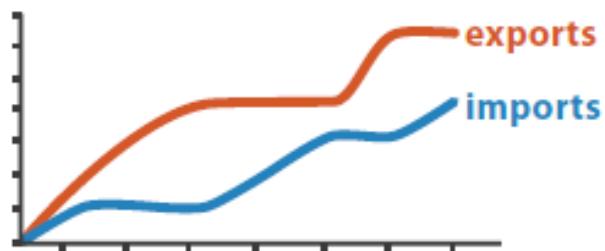
→ Derive



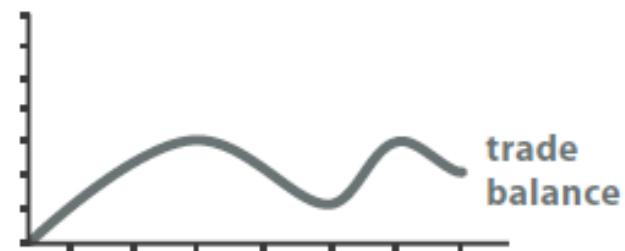
→ Search

e.g. the user could **annotate** the points within a cluster with a text label.

One interesting example of a record goal is to assemble a **graphical history**.



Original Data

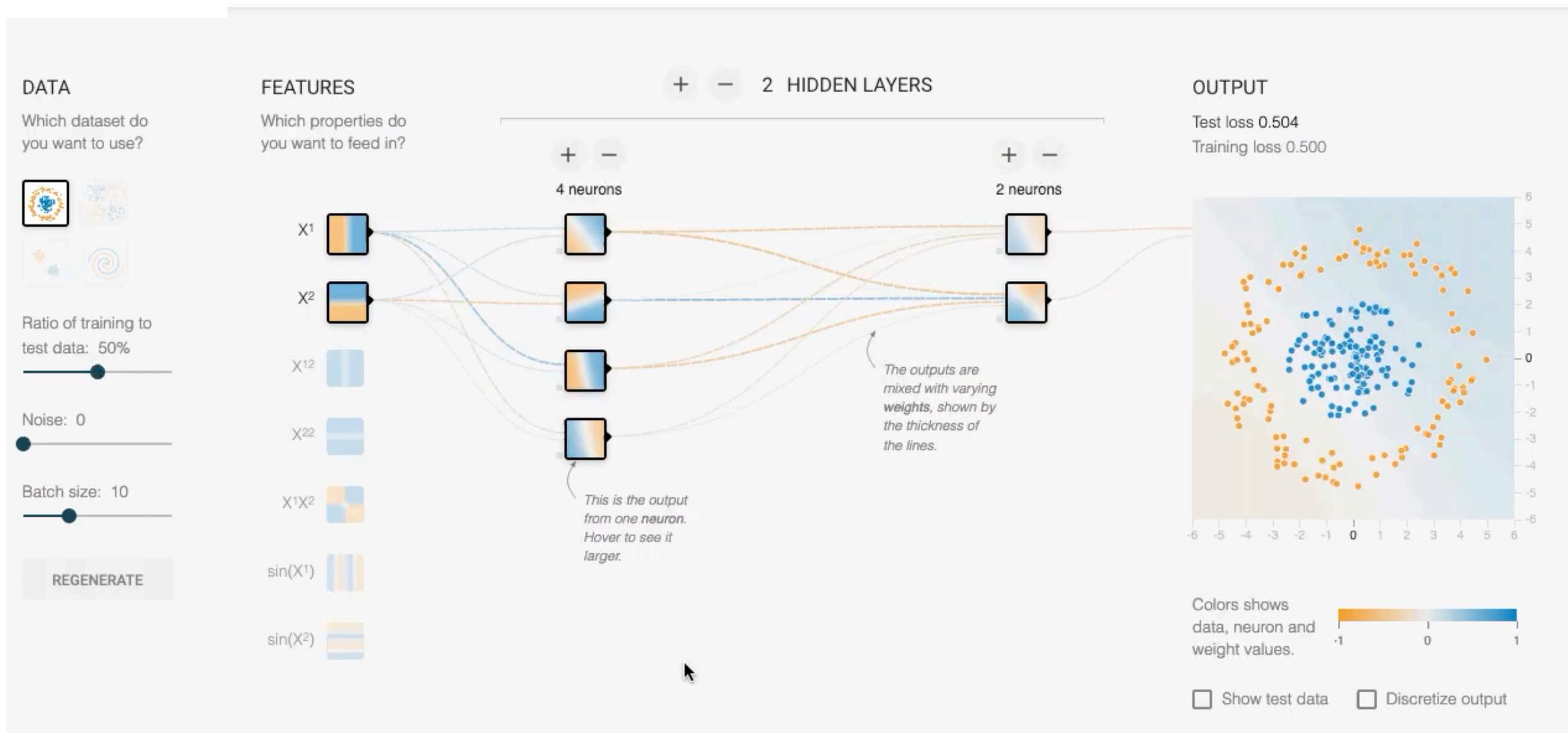


$$\text{trade balance} = \text{exports} - \text{imports}$$

Derived Data

# Produce

Epoch	Learning rate	Activation	Regularization	Regularization rate	Problem type
000,000	0.03	Tanh	None	0	Classification



# Enjoy

## NameVoyager: Explore baby names and name trends letter by letter

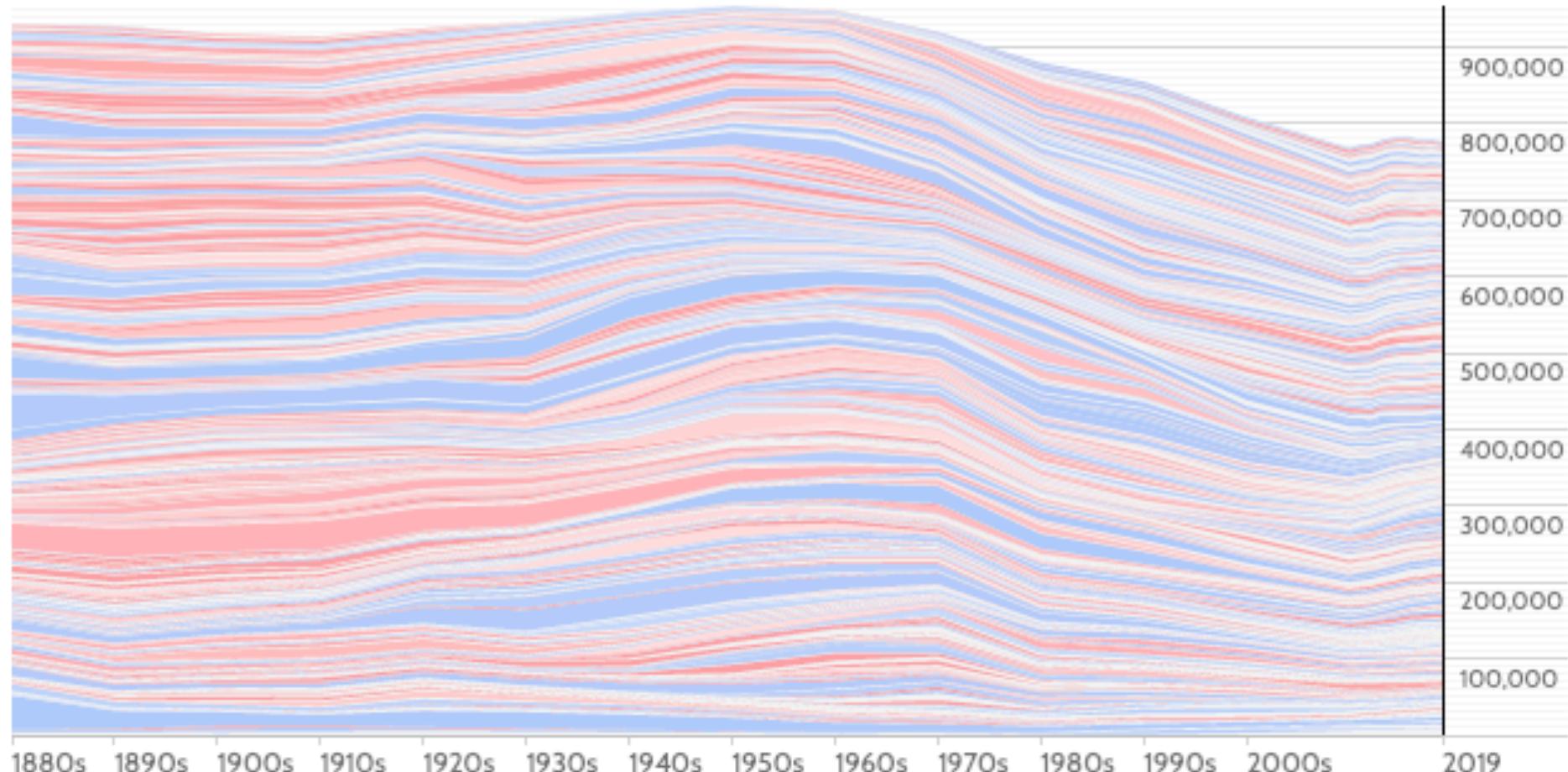
Baby Name >

Both  Boys  Girls

boys	1000	500	100	25	1
girls	1000	500	100	25	1

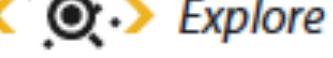
Current rank:

per million births



[babynamewizard](#)

## → Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

## → Query

→ Identify    → Compare    → Summarize



→ Identify    → Compare    → Summarize



Search/Find-> All Analyze cases require user search for items.

## Targets

→ All



Trends



Outliers



Features

→ Attributes

→ One



Distribution

→ Many

→ Dependency

→ Correlation

→ Similarity



Dependency



Correlation



Similarity

→ Paths



Topology



Paths



Spatial Data



Shape

→ What?

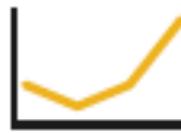
→ Why?

**Target** -> aspect of the data that is of **interest** to the user.

## Targets

### All Data

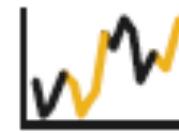
→ Trends



→ Outliers



→ Features

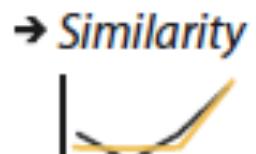
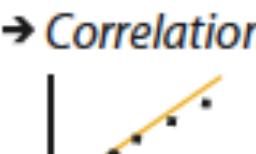
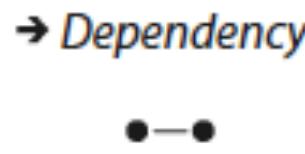


### Attributes

→ One

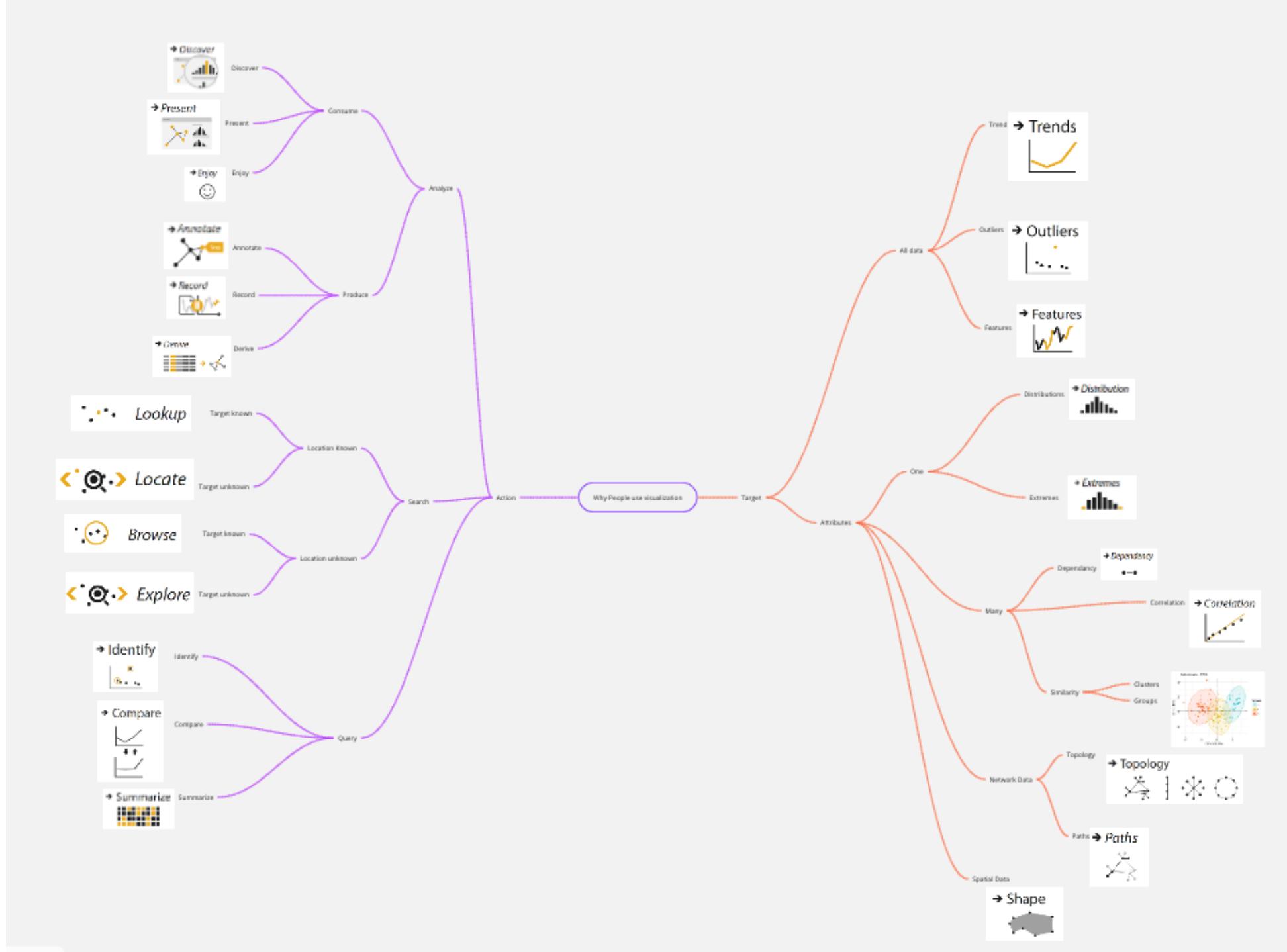


→ Many



→ Extremes



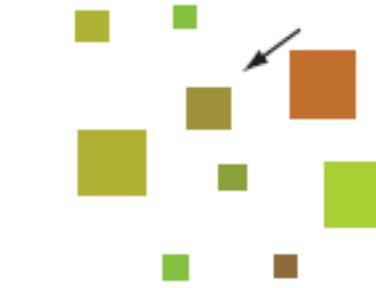
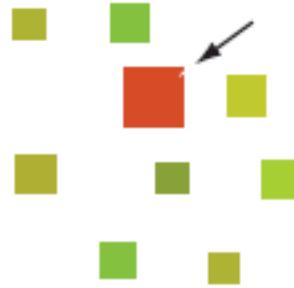


## Instructions

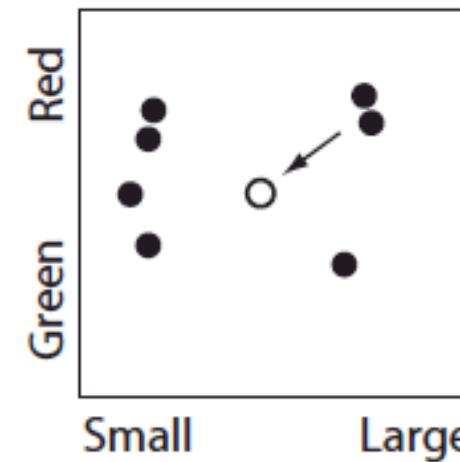
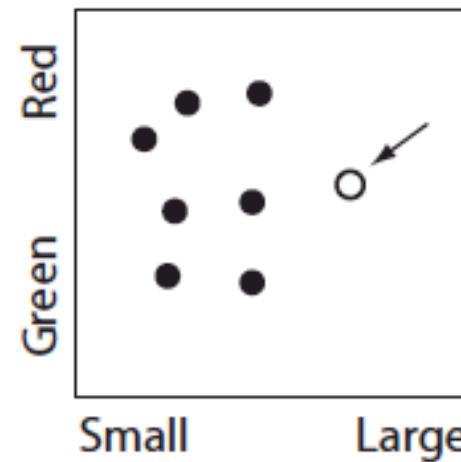
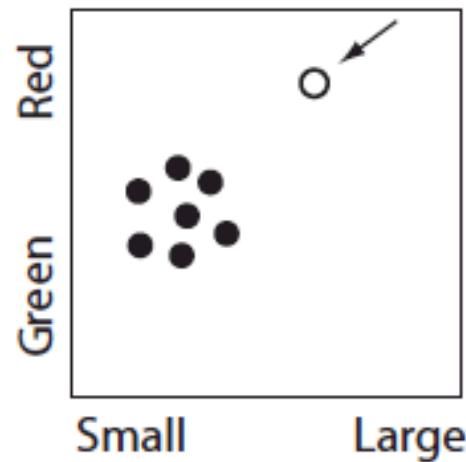
**Count how many times the  
players wearing white pass  
the basketball.**

Trying to find a **target** based on two features is called a **visual conjunctive search**

Objects to be searched



Corresponding feature space diagrams





difficult



easy

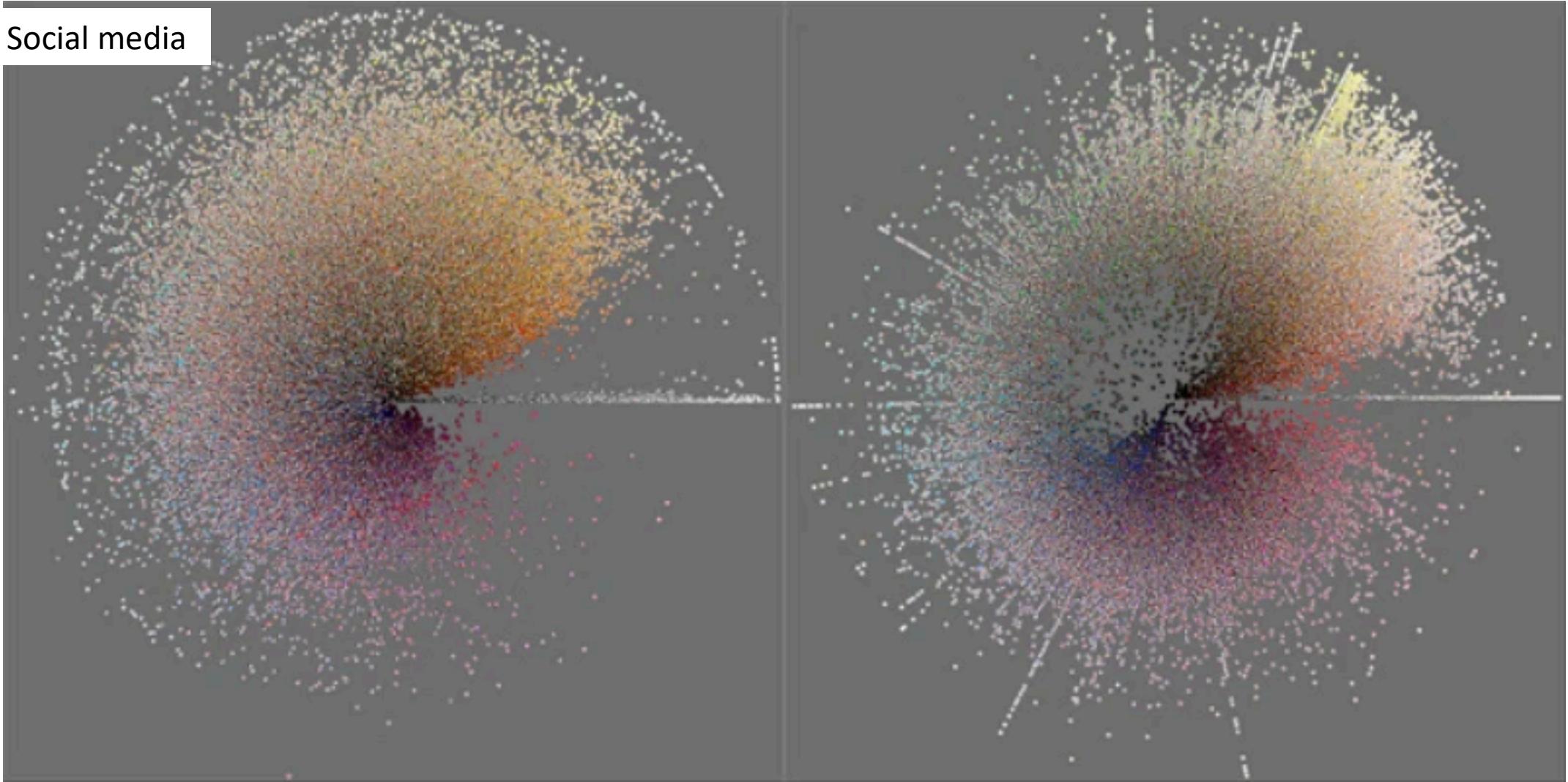


# Projects

# Examples of Modern Visualizations

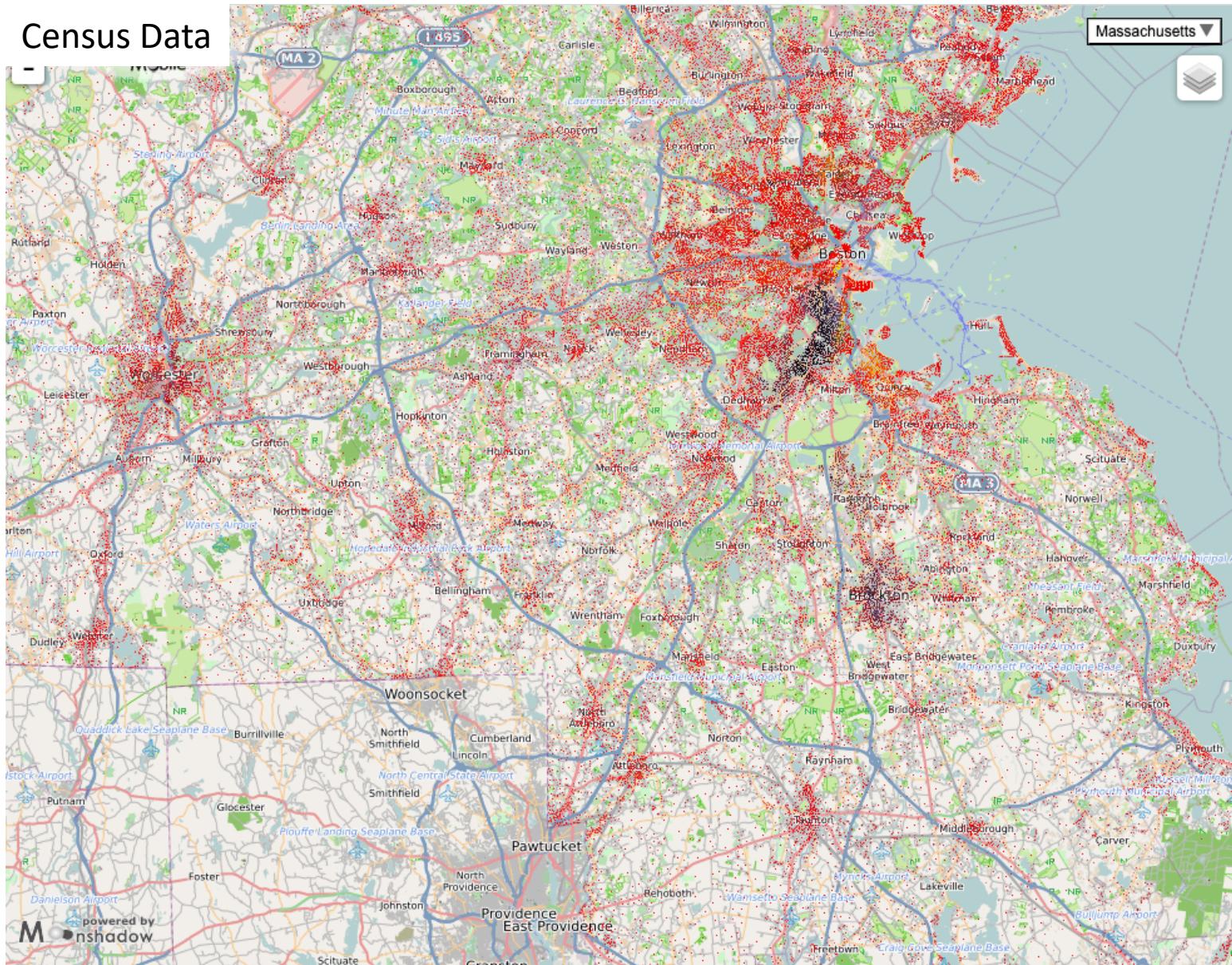
- Social media
  - Mobility for city planners
  - Community for Marketing
- Events and Census Data
  - Social interactions
  - Long term planning
  - Urban design
- Covid-19 (governmental actions)
- Simulation of Models (e.g. compartment model)
- Climate/weather stat (global warming, droughts, extreme events, and El Nino)
- Movie/Tv Series databases (Media)
- Traffic (cars, trains, airlines)
- Real estate (+ census data + climate change)

Social media



Visual signatures of San Francisco (left) compared with Bangkok (right). Based on 50,000 Instagram photos organized by brightness median (radius) and hue median (perimeter).

## Census Data



# CensusViewer

Login   Subscribe   Compare   Free

## Links & Information

[User manual](#) [Tutorials](#) [Give feedback](#) [FAQ](#) [Mailing list sign-up](#)

**Map Controls**

- Boundaries
- Shapes

**Universes**

- Universes

**Data Sources**

- MA 2014 Adults
- ▼ **Census 2010 Race**
  - Age
  - Gender
  - Race
  - Congressional Districts (111th)
  - County
  - State House Districts
  - State Senate Districts
  - Total
  - Heat Map
- ▼ **Census 2010 Latino**
  - Age
  - Gender
  - Latino
  - Congressional Districts (111th)
  - County
  - State House Districts
  - State Senate Districts
  - Total
  - Heat Map
- ▼ **Census 2000 Race**
  - Age

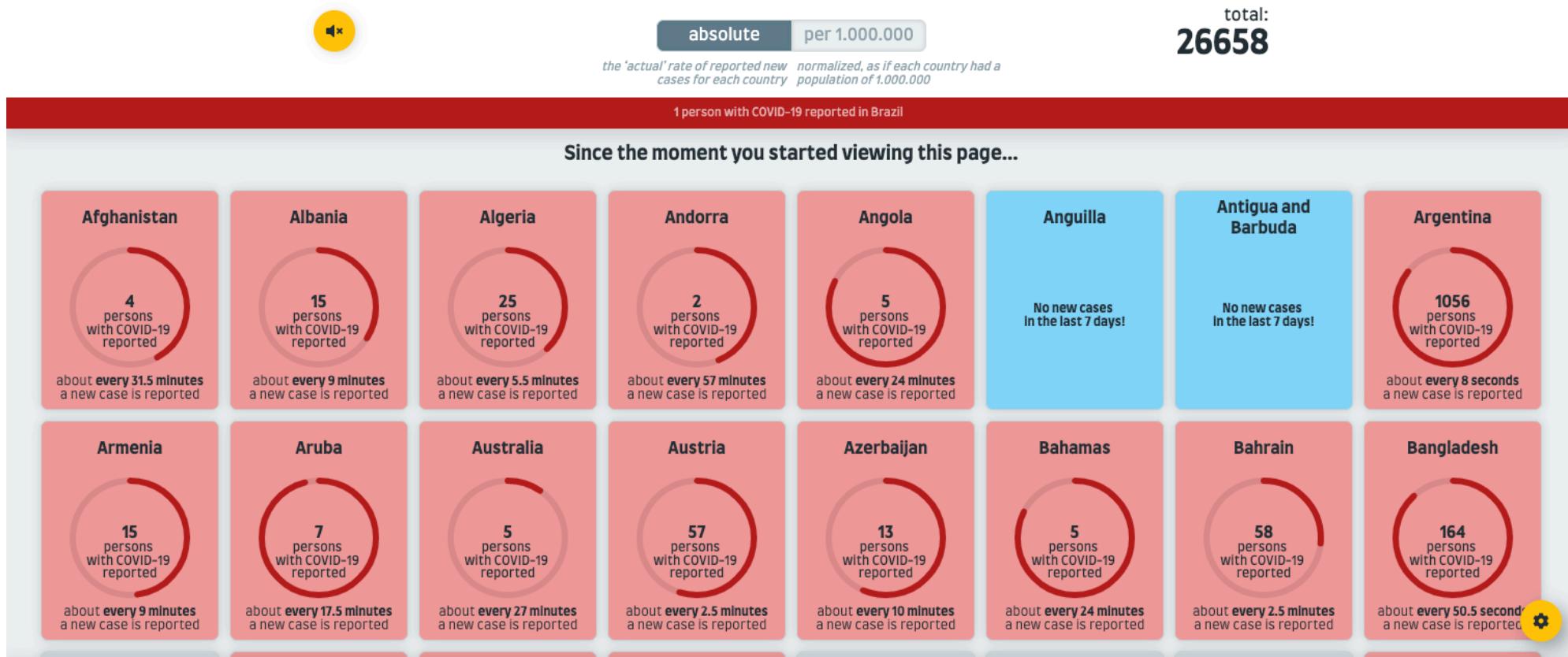
<http://massachusetts.us.censusviewer.com/client>

# Covid-19 Spread rate

<https://covidspreadingrates.org/>

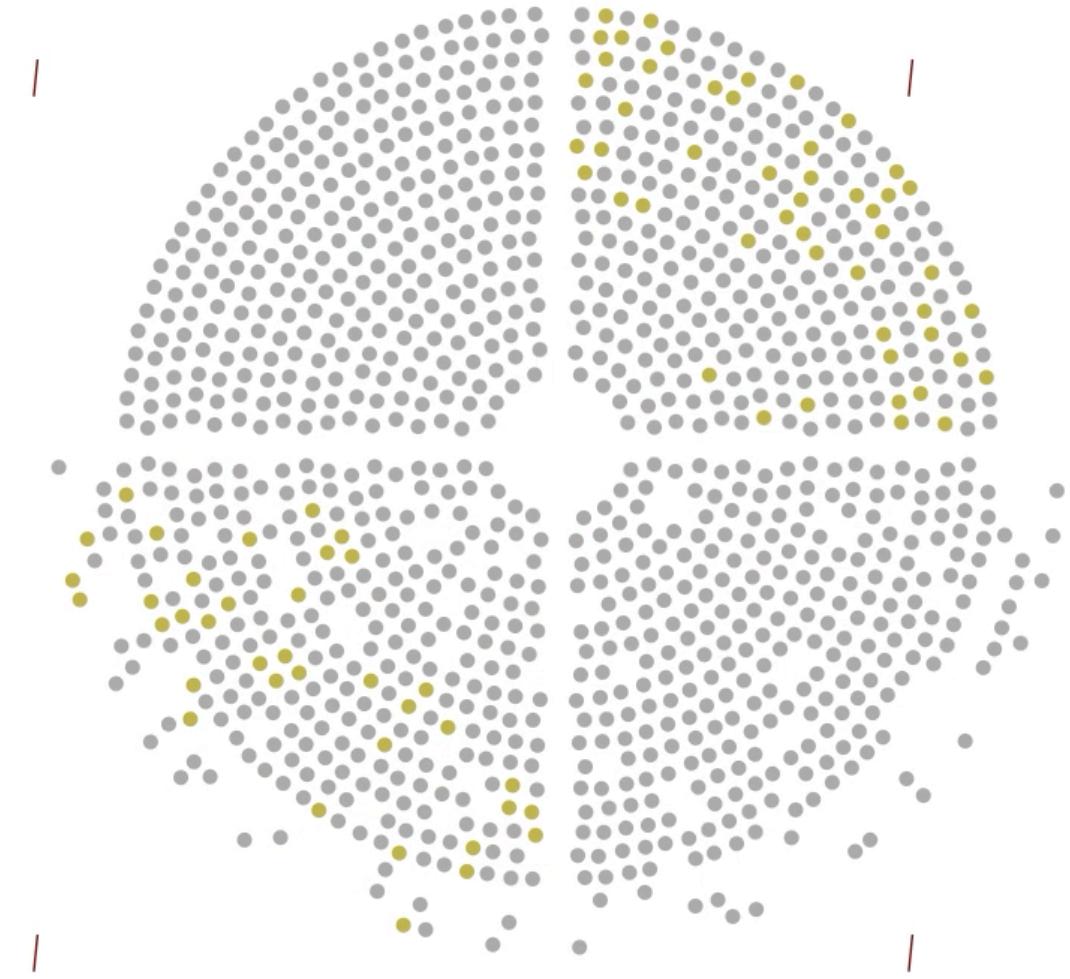
## COVID-19 spreading rates

Get an intuitive 'feel' of new COVID-19 cases spreading rates. This simulation shows you the *average* rate of newly reported COVID-19 cases between September 7, 2020 and September 14, 2020 for each country.



# COVID-19 spread sketch

<https://observablehq.com/@chromakode/xkcd-covid-spread-sketch?>



# Mobility Trend (Google Mobility data)

<https://ourworldindata.org/covid-mobility-trends>

How did the number of visitors change since the beginning of the pandemic?,  
United Kingdom, Feb 17, 2020

Our World  
in Data

The data shows how visitors to (or time spent in) categorized places change compared to baseline days – the median value from the 5-week period from January 3rd to February 6th 2020. This index is smoothed to the rolling 7-day average.



Source: Google COVID-19 Community Mobility Trends – Last updated 14 September, 19:01 (London time)

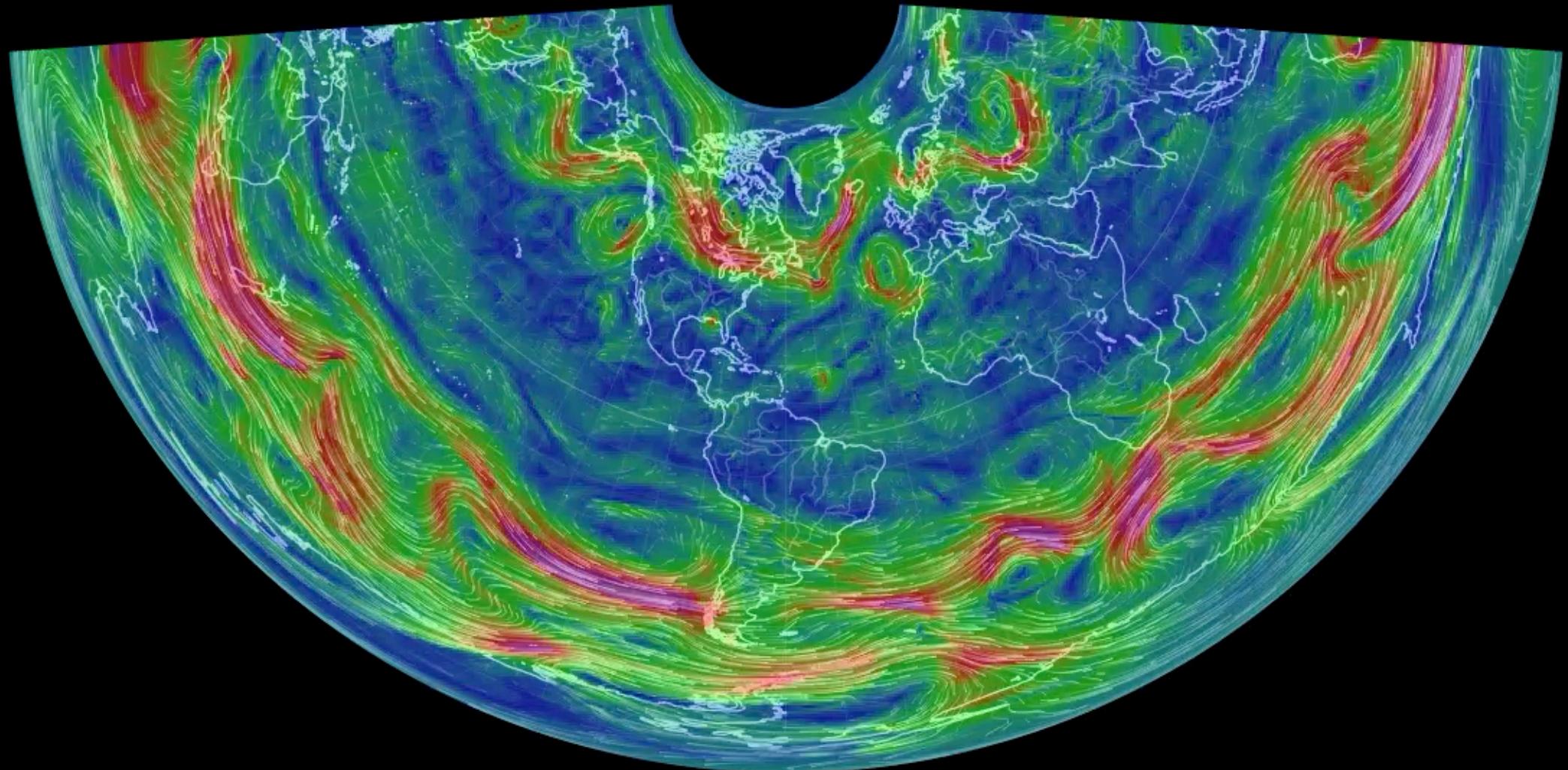
Note: It's not recommended to compare levels across countries; local differences in categories could be misleading.

CC BY

► Feb 17, 2020

Sep 11, 2020

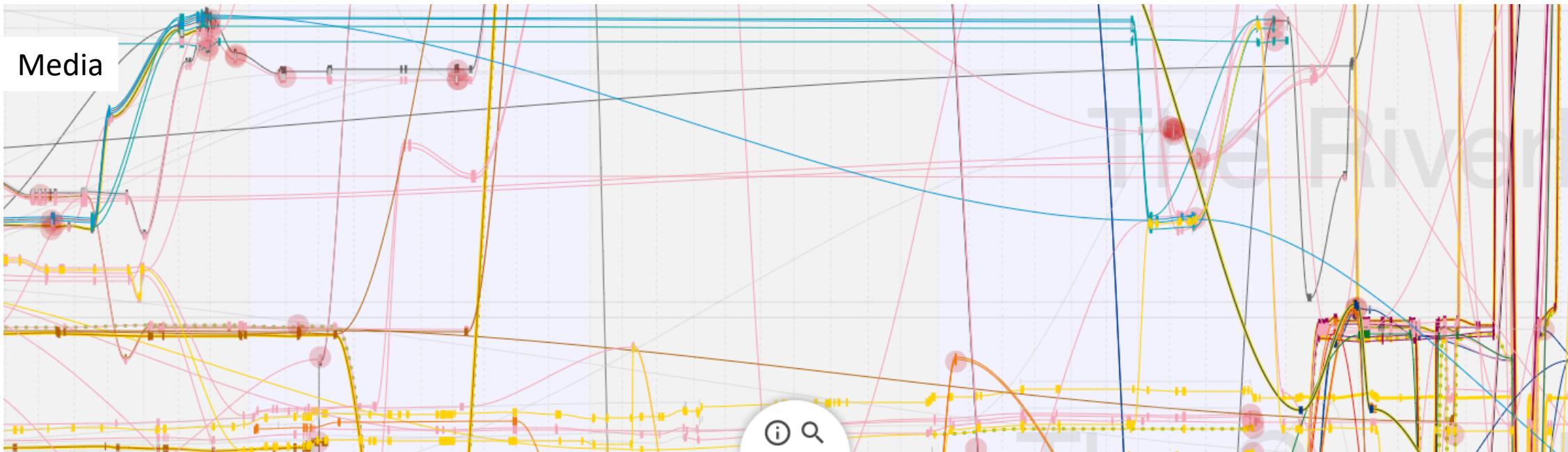
## Climate/Weather



earth

community  
EarthWindMap

[earth.nullschool.net](http://earth.nullschool.net)



## Key

- Stark + King
- Targaryen + Khaleesi
- Baratheon
- Lannister + Hand
- Night's Watch
- Dothraki + Khal
- Greyjoy
- Tyrell + Dead
- Wildlings
- Martell
- Frey + Dead
- Tully
- White Walkers
- Other

## Select

### Gender

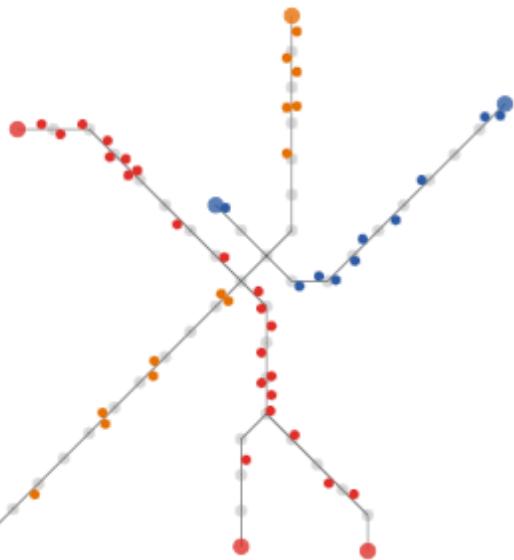
Male  Female  All Characters

### House

- ✓ Select a House
- Stark
  - Targaryen
  - Baratheon
  - Lannister
  - Night's Watch
  - Dothraki
  - Greyjoy
  - Tyrell
  - Wildlings
  - Martell
  - Frey

<https://jeffreylancaster.github.io/game-of-thrones/map/>

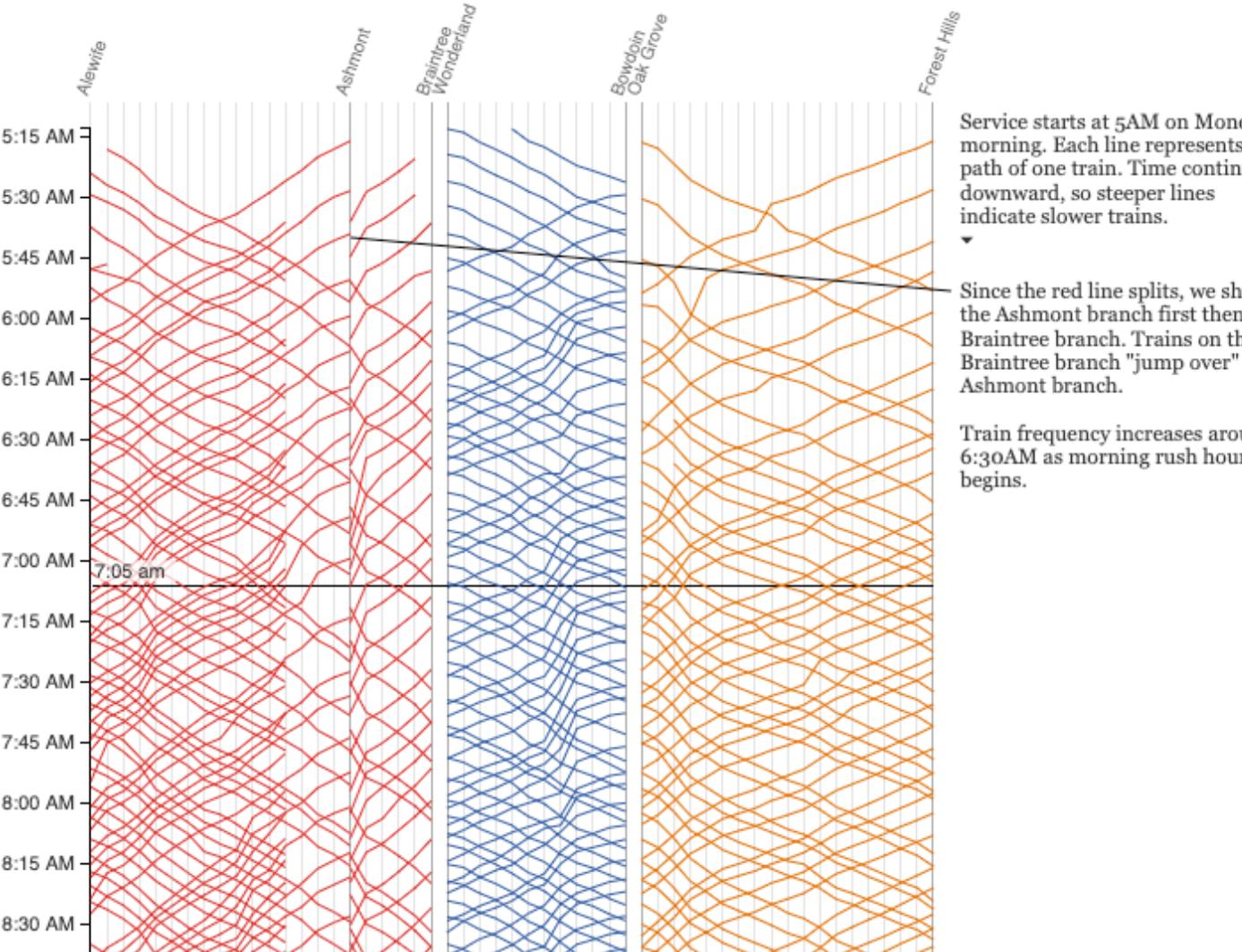
## Traffic



Locations of each train on the [red](#), [blue](#), and [orange](#) lines at 7:05 am. Hover over the diagram to the right to display trains at a different time.

Trains are on the right side of the track relative to the direction they are moving.

See the [morning rush-hour](#), [midday lull](#), [afternoon rush-hour](#), and the [evening lull](#).



# Solution of Folium Choropleth

1. Get FIPS Shape files from US Census website
2. Convert shape file into GeoJSON
3. Import Pandas and Folium
4. Clean NYT Covid19 Data (assign an arbitrary FIPS code for NYC)
5. Create folium Map object
6. Create Choropleth Object
7. Set Data and GeoJSON path.
8. Done!

# Cool Infographics

[https://umassd.primo.exlibrisgroup.com/permalink/01MAA\\_DM\\_INST/kde4of/alma9913799190001301](https://umassd.primo.exlibrisgroup.com/permalink/01MAA_DM_INST/kde4of/alma9913799190001301)

# Cool Infographics

Effective Communication with  
Data Visualization and Design

By  
Randy Krum



WILEY

# Ten guidelines for effective data visualization in scientific publications

[https://umassd.primo.exlibrisgroup.com/permalink/01INST1f5rao8/cdi\\_crossref\\_primary\\_10\\_1016\\_j\\_envsc12\\_006](https://umassd.primo.exlibrisgroup.com/permalink/01INST1f5rao8/cdi_crossref_primary_10_1016_j_envsc12_006)

Value encoding attribute			
	Length	Width	Orientation
Form			/
	Size	Shape	Curvature
	● ● ● ● ● ● ● ● ● ● ● ●	   ■     	) ) ) ) ) ) ) ) ) ( ) )
Color	Enclosure	Blur	
	██████████ ██████████ ███████ ███	● ● ● ● ● ■ ● ● ● ● ● ●	
Spatial Position	Hue	Intensity	Transparency
	● ● ● ● ● ● ● ● ● ● ● ●	● ● ● ● ● ● ● ● ● ■ ● ●	● ● ● ● ● ● ● ● ●
Motion	2-D Position	Spatial Grouping	Density
	● ● ● ●	● ● ● ● ● ● ● ● ●	○ ○ ○ ○ ○ ○
Direction		Pathway	
			